DISTRIBUZIONI DI FREQUENZA

corso di statistica

Francesco Lagona

Università Roma Tre

tabella unità-variabili

	variabili					
unità	X_1	X_2	$\dots$	X_k	$\dots$	X_K
1	x_{i1}	x_{12}	$\dots$	x_{1k}	$\dots$	x_{1K}
2	x_{21}	x_{22}	$\dots$	x_{2k}	$\dots$	x_{2K}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	x_{i1}	x_{i2}	$\dots$	x_{ik}	$\dots$	x_{iK}
$\vdots$						
n	x_{n1}	x_{n2}	$\dots$	x_{nk}	$\dots$	x_{nK}

- riga i: profilo dell'unità i
- colonna k: distribuzione unitaria della variabile X_k

matrice unità-variabili

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2K} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{i1} & x_{i2} & \dots & x_{ik} & \dots & x_{iK} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} & \dots & x_{nK} \end{pmatrix}$$

- profilo dell'unità i: $(x_{i1}, x_{i2} \dots x_{ik} \dots x_{iK})$
- distribuzione unitaria della variabile X_k: $(x_{1k}, x_{2k} \dots x_{ik} \dots x_{nk})$

distribuzioni unitarie univariate

- concentriamo l'attenzione su una singola distribuzione univariata: $(x_1 \dots x_i \dots x_n)$
- esempi:
  - distribuzione unitaria del genere (qualitativa sconnessa con supporto $\{M, F\}$):

$$(M, F, M, F, F, F, M, F, F, M)$$

  - distribuzioe unitaria della residenza (qualitativa sconnessa con supporto $\{RM, VT, LT, MI\}$):

$$(RM, VT, LT, LT, MI, RM, RM)$$

  - distribuzione unitaria del giudizio su un prodotto (qualitativa ordinabile con supporto $\{scarso, suff, discreto, buono\} = \{S, SS, D, B\}$)

$$(S, SS, SS, D, S, B, D, B, S, SS, SS)$$

  - distribuzione unitaria del numero di incidenti accaduti giornalmente in un mese (quantitativa discreta con supporto {0,1,2,3,4,5,6})

$$(3, 3, 2, 3, 1, 1, 2, 1, 2, 2, 1, 0, 2, 6, 2, 2, 1, 0, 3, 1, 3, 4, 1, 3, 4, 3, 2, 0, 0, 4)$$

distribuzioni di frequenza univariate

$(M, F, M, F, F, F, M, F, F, M) \quad \Rightarrow$

x_k	n_k
M	4
F	6
	10

$(RM, VT, LT, LT, MI, RM, RM) \quad \Rightarrow$

x_k	n_k
RM	3
VT	1
LT	2
MI	1
	7

distribuzioni di frequenza univariate

$$(S, SS, SS, D, S, B, D, B, S, SS, SS) \quad \Rightarrow$$

x_k	n_k
S	3
SS	4
D	2
B	2
	11

distribuzioni di frequenza univariate

$$(3, 3, 2, 3, 1, 1, 2, 1, 2, 2, 1, 0, 2, 6, 2, 2, 1, 0, 3, 1, 3, 4, 1, 3, 4, 3, 2, 0, 0, 4)$$

x_k	n_k
0	4
1	7
2	8
3	7
4	3
6	1
	30

distribuzioni di frequenza univariate

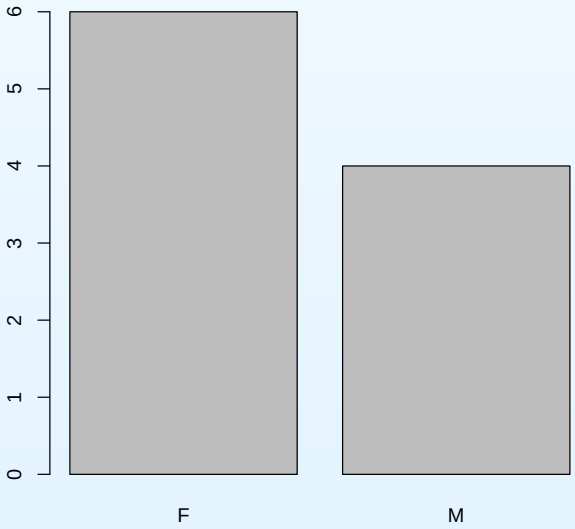
- in generale:

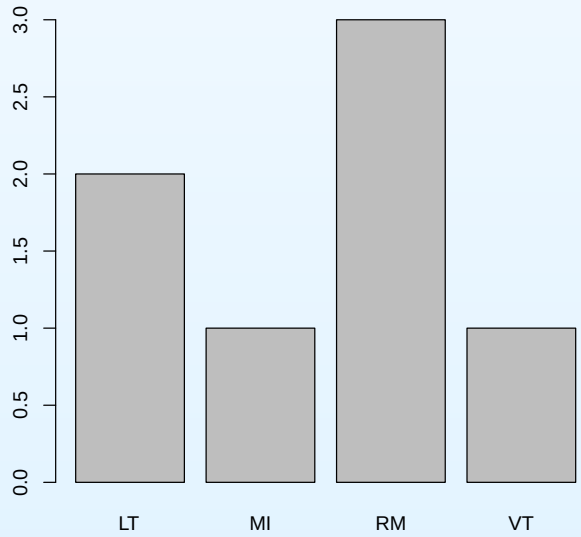
X	$n(x)$
x_1	n_1
x_2	n_2
$\vdots$	$\vdots$
x_k	n_k
$\vdots$	$\vdots$
x_K	n_K
	n

diagrammi a barra

x_k	n_k
M	4
F	6
	10

x_k	n_k
RM	3
VT	1
LT	2
MI	1
	7

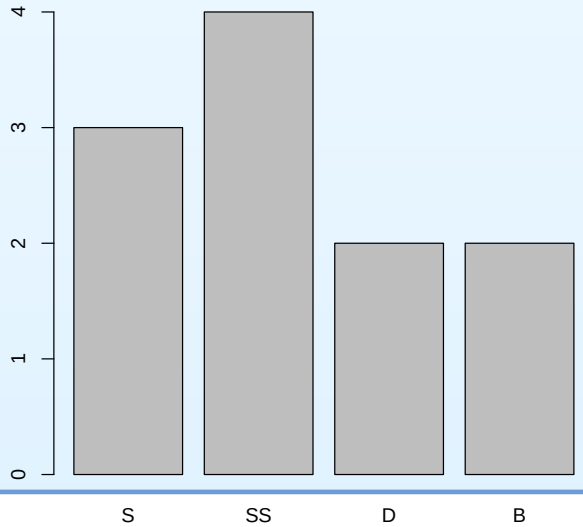


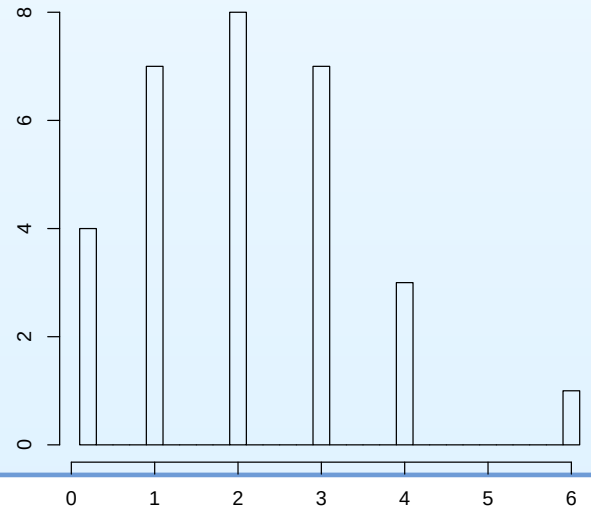


diagrammi a barra

x_k	n_k
S	3
SS	4
D	2
B	2
	11

x_k	n_k
0	4
1	7
2	8
3	7
4	3
6	1
	30





il caso delle variabili continue

- esempio: differenza (in minuti) tra il tempo di arrivo effettivo e quello dichiarato dall'ente ferroviario di $n = 50$ treni ad una stazione:

3.91	0.14	1.76	0.66	-1.06	-3.79	3.34	-0.96	-2.85
-0.11	-5.82	-1.37	-3.27	-0.14	5.04	-4.43	-4.58	0.33
1.73	-8.39	4.19	-1.44	-3.72	-3.44	-0.25	-5.46	6.85
-2.79	-4.77	-5.42	3.22	6.25	0.32	-8.94	-3.88	-0.47
0.04	1.41	3.19	0.65	4.60	5.83	-3.26	1.89	-1.77

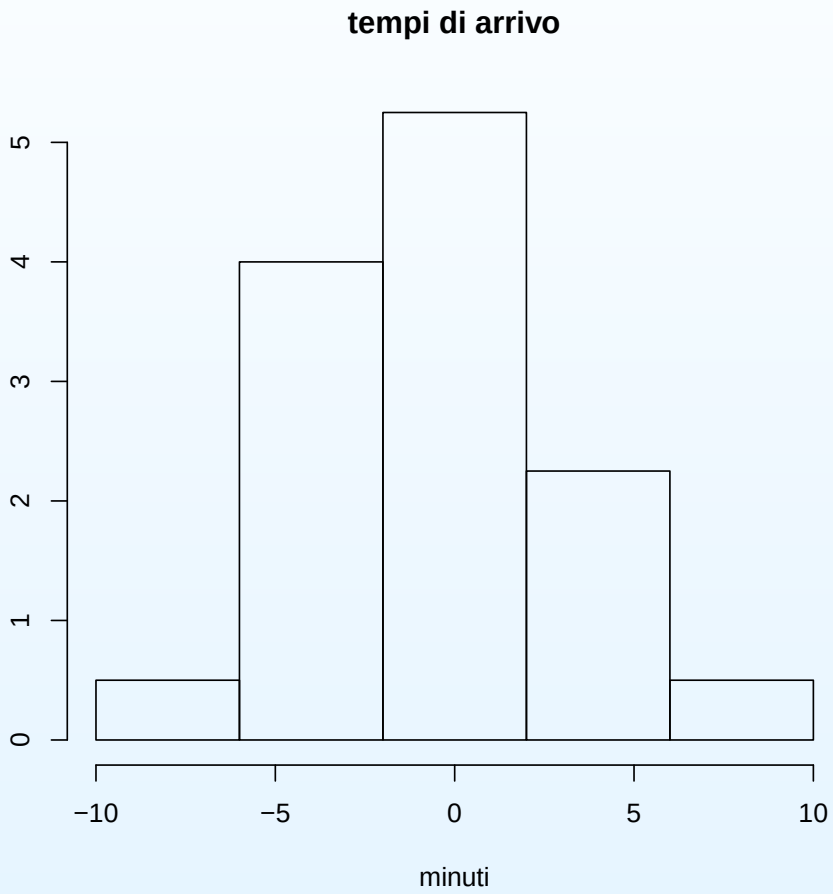
- la distribuzione di frequenze non sintetizza quella unitaria

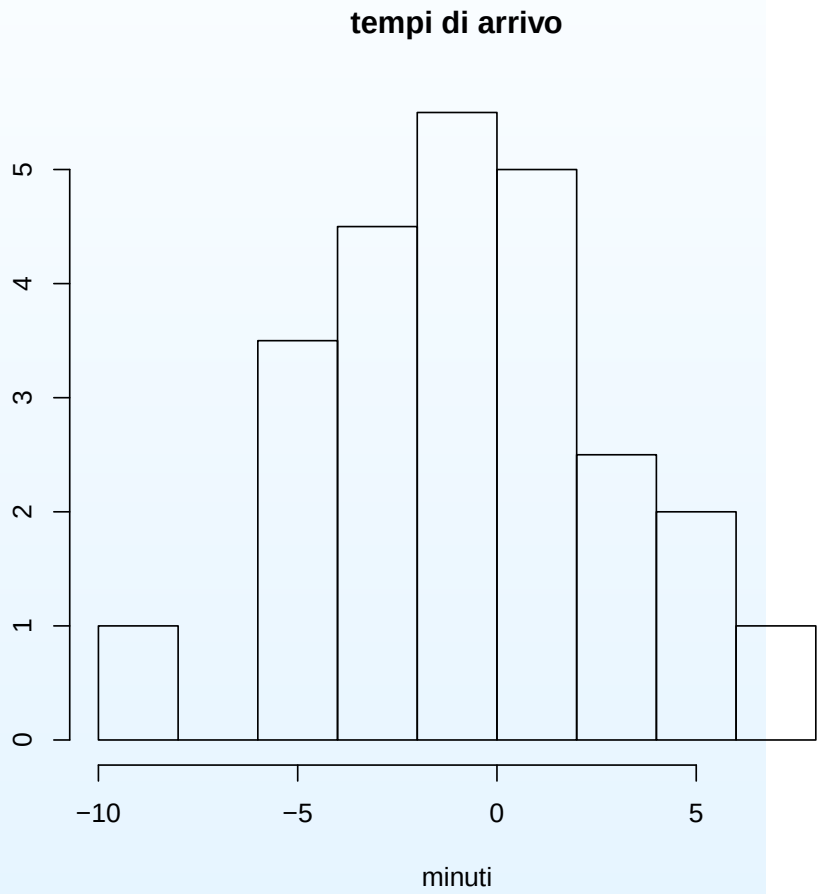
distribuzioni di frequenze

$classe$	$frequenza$
$[-10, -6)$	2
$[-6, -2)$	16
$[-2, 2)$	21
$[2, 6)$	9
$[6, 10]$	2

$classe$	$frequenza$
$[-10, -8)$	2
$[-8, -6)$	0
$[-6, -4)$	7
$[-4, -2)$	9
$[-2, 0)$	11
$[0, 2)$	10
$[2, 4)$	5
$[4, 6]$	4
$[6, 8]$	2

istogrammi


tempi di arrivo
0
1
2
3
4
5
−10
−5
0
5
10
minuti


tempi di arrivo
0
1
2
3
4
5
−10
−5
0
5
minuti